

Lead Behaviour in Bitcoin Markets

Ying Chen ¹, Paolo Giudici ^{2,*} , Branka Hadji Misheva ³ and Simon Trimborn ⁴ 

¹ Department of Mathematics and Risk Management Institute, National University of Singapore, Singapore 119077, Singapore; matcheny@nus.edu.sg

² Department of Economics and Management, University of Pavia, 27100 Pavia, Italy

³ School of Engineering, ZHAW University of applied sciences, 8005 Zurich, Switzerland; hadji@zhaw.ch

⁴ Department of Mathematics, National University of Singapore, Singapore 119077, Singapore; simon.trimborn@nus.edu.sg

* Correspondence: giudici@unipv.it

Received: 5 November 2019; Accepted: 31 December 2019; Published: 4 January 2020



Abstract: We aim to understand the dynamics of Bitcoin blockchain trading volumes and, specifically, how different trading groups, in different geographic areas, interact with each other. To achieve this aim, we propose an extended Vector Autoregressive model, aimed at explaining the evolution of trading volumes, both in time and in space. The extension is based on network models, which improve pure autoregressive models, introducing a contemporaneous contagion component that describes contagion effects between trading volumes. Our empirical findings show that transactions activities in bitcoins is dominated by groups of network participants in Europe and in the United States, consistent with the expectation that market interactions primarily take place in developed economies.

Keywords: bitcoin markets; bitcoin trading volumes; network models

1. Introduction

The bitcoin is the leading cryptocurrency by capitalisation, with a market share greater than 50% of the total cryptocurrency market, corresponding to 330 billion USD at its historical peak, in December 2017. Recent studies report that the same market capitalisation is concentrated on a limited number of owners. In particular, Credit Swiss in January 2018 provided a study which indicates that 97% of Bitcoins are held by 4% of all Bitcoin addresses. Bloomberg reported similar findings by suggesting that about 40 percent of Bitcoin is held by perhaps 1000 users.

The previous empirical findings suggest that the trading movement by a few bitcoin owners has the potential to cause major disruptions in the price of all cryptocurrencies. An example of this is the transaction that took place on 12 November 2017, when a user moved 25,000 Bitcoins, worth at the time USD159 million, to an exchange. A very important research question is therefore: “to find the bitcoin owners who are most connected in the markets, in terms of trading volumes”.

Unfortunately, the anonymity of bitcoin transactions makes very difficult to find an answer to the previous question. However, although it may be difficult to trace the “physical” identity of the users, it may be possible to understand their “statistical” identity, applying appropriate econometric models to the (very large) database of payments generated by bitcoin trades themselves. This may help to answer a less demanding, but still important research question: “to find groups of bitcoin owners who are most connected in the market, in terms of trading volumes”.

In this study, we classify bitcoin owners according to their observed trading behaviour, in ten classes of increasing average size. We add to this classification the geographical area of the owners, defined (very broadly) by the continent to which they belong. We then apply network econometric models to understand the map of interconnections that exist between the defined owner groups and, in this way, identify the trading groups who lead bitcoin markets, along time.

The econometric research on the dynamics of cryptocurrency markets has mainly been focused on the issue of price discovery and prediction. In this context, many of the stylized facts that are valid for traditional financial time series apply, to some extent, also in the context of these alternative currencies [Elendner et al. \(2017\)](#). A large stream of papers consider the dynamics of crypto prices, using VAR models ([Bianchi \(2019\)](#); [Catania et al. \(2019\)](#); [Bohte and Rossini \(2019\)](#); [Giudici and Abu-Hashish \(2019\)](#)), VECM models ([Giudici and Pagnottoni \(2019a\)](#), [\(2019b\)](#)), similarity networks [Giudici and Polinesi \(2019\)](#) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models [Bouoiyour et al. \(2016\)](#). The results from the different papers, however, seem far from consistent. In our view, this is mostly due to the nature of the cryptocurrencies. For example, they are much more volatile compared to traditional currencies, their exchange rates cannot be assumed to be independently and identically distributed and their global nature limits researchers' ability to account for systematic causal factors.

In our opinion, it becomes necessary to move away from traditional price volatility models, and focusing on the identification of the mechanisms that drive trading behaviour, as in our research question. The available literature on trading volume dependency in cryptocurrency markets is very limited. Notable exception to this are the papers by [Tasca et al. \(2018\)](#), [Foley et al. \(2019\)](#) and [Chen et al. \(2018\)](#). In particular, [Tasca et al. \(2018\)](#) attempt to identify different clusters within the Bitcoin economy by analyzing the trading patterns and ascribing them to particular business categories. Using network-based methods, the authors have identified three market regimes that have characterized Bitcoin transactions.

Our work intends to extract the network of payment relationship between Bitcoin users, owners, similar to [Tasca et al. \(2018\)](#). We extend their work, acquiring evidence on whether trading volumes behaviors of different groups of Bitcoin traders, defined by volume size and geographical region, are interconnected and, therefore, affect each other.

From an econometric viewpoint, we propose an econometric network model which extends Vector Autoregressive models. The extension is based on network models, which improve over pure autoregressive models, as they introduce a contemporaneous contagion component that describes contagion effects between groups of traders.

The validity of the model was demonstrated in recent studies on systemic risk, in which researchers have proposed correlation network models, able to combine the rich structure of financial networks (see, e.g., [Lorenz et al. \(2009\)](#); [Battiston et al. \(2012\)](#)) with a more parsimonious approach that can estimate contagion effects from the dependence structure among market prices. The first contributions in this framework are [Billio et al. \(2012\)](#) and [Diebold and Yilmaz \(2014\)](#), who derive contagion measures based on Granger-causality tests and variance decompositions. More recently, [Ahelegbey et al. \(2016\)](#) and [Giudici and Spelta \(2016\)](#) have extended this methodology introducing stochastic correlation networks.

While bivariate systemic risk models (such as [Acharya et al. \(2012\)](#), [Acharya et al. \(2016\)](#) and [Adrian and Brunnermeier \(2015\)](#)) explain whether the risk of an institution is affected by a market crisis event or by a set of exogenous risk factors, correlation network models explain whether the same risk depends on contagion effects, in a cross-sectional perspective.

We extend the approach of [Giudici and Spelta \(2016\)](#) enriching their graphical Gaussian model with an autoregressive component derived through a VAR model, as in [Ahelegbey et al. \(2016\)](#). In contrast with the latter, we employ partial correlations rather than correlations, and we do not follow a Bayesian approach.

We remark that our work is related to some recent papers that explore the cross-country trading in cryptocurrency markets [Makarov and Schoar \(2019\)](#), the network dynamics across cryptocurrency markets [Ji et al. \(2019\)](#) and the information content of trading volumes in crypto investing [Bianchi \(2019\)](#); [Bouri et al. \(2019\)](#). We combine the views of the previous paper into a network-based analysis of bitcoin trading patterns across countries and trading groups.

To demonstrate our methodology, we will consider the all world's bitcoin transactions, independently of the exchange in which they were traded, in the time period 25 February 2012 to 17 July 2017.

Our empirical findings show that transactions activities in bitcoins is dominated by groups of network participants in Europe and in the United States, consistent with the conventional wisdom that posits market interactions, at least nominally, primarily take place in developed economies.

The paper is organized as follows: Section 2 contains our proposed model; Section 3 presents the available data; Section 4 the empirical application of the proposed model to the obtained data; Section 5 contains some concluding remarks.

2. Proposal

Let y_t^i be the traded volume of Bitcoin by a specific group of traders i ($i = 1, \dots, I$), at time t ($t = 1, \dots, T$). We assume that y_t^i is a function of: (a) an autoregressive element that captures the dependence on the past trading volumes of the same group; (b) a cross-sectional element that captures the contemporaneous dependence on the trading volumes of other groups; (c) a stochastic residual. Mathematically, we assume that in the case of the Bitcoin traded volumes, for each volume i and time t the following equation holds:

$$y_t^i = \sum_{p=1}^{p_0} \alpha_p^i y_{t-p}^i + \sum_{j \neq i} \beta^{ij} y_t^j + \epsilon_t^i, \quad (1)$$

where p is a time lag (with a maximum value of $p_0 < t$), α_p^i and β^{ij} are the coefficients which are to be estimated, and ϵ_t^i are residuals, which we assume standard Gaussian and independent.

Equation (1) models the Bitcoin volume dynamics as a structural VAR, in which the traded volume in each group depends on its p past values, through the idiosyncratic autoregressive component $\sum_{p=1}^{p_0} \alpha_p^i y_{t-p}^i$ and, in addition, it depends on the contemporaneous values of the other groups, through the systemic component $\sum_{j \neq i} \beta^{ij} y_t^j$.

Defining B_0 as a $I \times I$ symmetric matrix with null diagonal elements containing the contemporaneous coefficients, the previous model can be expressed in a more compact matrix form, as follows:

$$Y_t = \sum_{p=1}^{p_0} A_p Y_{t-p} + B_0 Y_t + \epsilon_t, \quad (2)$$

where Y_t is a I -dimensional vector containing the traded volumes of all groups at time t , Y_{t-p} is the same vector, lagged at time $t - p$, A_p is a $I \times I$ matrix that contains the autoregressive coefficients and ϵ_t is a vector of residuals.

In the following step, we transform the model in (2) into a reduced form for the purpose of facilitating the estimation process, thus becoming:

$$Y_t = \Gamma_1 Y_{t-1} + \dots + \Gamma_{p_0} Y_{t-p_0} + U_t, \quad (3)$$

with

$$\begin{cases} \Gamma_1 = (\mathbb{I} - B_0)^{-1} A_1, \\ \dots \\ \Gamma_{p_0} = (\mathbb{I} - B_0)^{-1} A_{p_0}, \\ U_t = (\mathbb{I} - B_0)^{-1} \epsilon_t. \end{cases} \quad (4)$$

This reduced form allows the estimation of the vectors of modified autoregressive coefficients $\Gamma_1, \dots, \Gamma_{p0}$, using time series data on the traded volumes contained in the stacked vector $\{Y_1, \dots, Y_t, \dots, Y_T\}$.

However, we are not interested in estimating Γ_p . In fact, the purpose of this analysis is to disentangle its autoregressive and contemporaneous components, thus separately estimating $\{A_1, \dots, A_{p0}\}$ and B_0 . In this sense, once B_0 is obtained, $\{A_1, \dots, A_{p0}\}$ can be derived from (4).

To estimate B_0 , note that $(\mathbb{I} - B_0)U_t = \varepsilon_t$, so that $U_t = B_0U_t + \varepsilon_t$. This implies that, for each group i ,

$$U_t^i = \sum_{j \neq i} \beta^{ij} U_t^j + \varepsilon_t^i, \quad (5)$$

meaning that the off-diagonal elements of B_0 can be obtained regressing each modified residual, derived from the application of (3), on those of the other groups.

Please note that the regression model in (5) is based on the transformation derived in Equation (4), which makes the modified residuals correlated. The direction of such correlation is, however, unknown. In the application of (5) it is, therefore, not clear which volume residual assumes the form of a response variable, and which one of an explanatory regressor.

To determine the direction of such dependence, we propose to approximate each pair of regression coefficients β^{ij} and β^{ji} , with their partial correlation coefficient, which is undirected.

Mathematically, let $\Sigma = \text{Corr}(U)$ be the correlation matrix between the modified residuals, and let Σ^{-1} be its inverse, with elements σ^{ij} . The partial correlation coefficient $\rho_{ij|S}$ between the residuals U^i and U^j , conditional on the remaining residuals $(U^s, s = 1, \dots, S)$, where $S = I \setminus \{i, j\}$, can be obtained as:

$$\rho_{ij|S} = \frac{-\sigma^{ij}}{\sqrt{\sigma^{ii}\sigma^{jj}}}. \quad (6)$$

It can be shown that:

$$|\rho_{ij|S}| = \sqrt{\beta^{ij} \cdot \beta^{ji}}, \quad (7)$$

which means that the absolute value of the partial correlation coefficient between U^i and U^j , given all the other residuals, can be obtained as the geometric average between the coefficients β^{ij} and β^{ji} defined by equation (5) setting, respectively, i rather than j as response variables. Equation (7) justifies the replacement of β^{ij} and β^{ji} with their corresponding partial correlation coefficient $\rho_{ij|S}$.

From an economic viewpoint, the partial correlation coefficient expresses how the trading volume of node i is affected by the contemporaneous trading volume of node j ($j \neq i$), keeping the other volumes fixed.

An important advantage that derives from the employment of partial correlations lies in the possibility of employing correlation network models based on the conditional independence relationships described by partial correlations.

More precisely, let us assume that the vectors U_t are independently distributed according to a multivariate normal distribution $\mathcal{N}_I(0, \Sigma)$, where Σ represents the correlation matrix (that we assume to be non-singular).

A correlation network model can be represented by an undirected graph G such that $G = (V, E)$, with a set of nodes $V = \{1, \dots, I\}$, and an edge set $E = V \times V$ that describes the connections between the nodes. G can be represented by a binary adjacency matrix E with elements e_{ij} , each of them providing the information of whether a pair of vertices in G is (symmetrically) linked between each other ($e_{ij} = 1$) or not ($e_{ij} = 0$). If the nodes V of G are put in correspondence with the random variables U_1, \dots, U_I , the edge set E induces conditional independences on U via the so-called Markov properties (see e.g., Lauritzen (1996)).

Following up on (7), Whittaker (1990) proved that the following equivalence holds:

$$\rho_{ij|S} = 0 \iff U_i \perp U_j | U_{V \setminus \{i,j\}} \iff e_{ij} = 0 \quad (8)$$

where the symbol \perp indicates conditional independence.

From a graph theoretic viewpoint, the previous equivalence means that a link between two volume residuals is present if and only if the corresponding partial correlation coefficient is significantly different from zero.

From a financial viewpoint, the previous equivalence implies that, if the partial correlation between two measures is equal to zero, the corresponding volumes residuals are conditionally independent and, therefore, the corresponding groups do not (directly) impact each other.

From a statistical viewpoint, it is also possible to test the null hypotheses that two groups of Bitcoin owners are conditionally independent by controlling whether the corresponding partial correlation coefficient is equal to zero, by means of the statistical test described in Whittaker (1990).

However, this poses a problem of multiple testing, and correcting for this problem could result in loss of power (for example using Bonferroni's inequality). One of the most widely used method for limiting the number of spurious edges—while at the same time obtaining networks that are more interpretable,—is through the use of a regularization approach. One such prominent approach of regularization is the 'least absolute shrinkage and selection operator (LASSO) which in its essence, allows us to set estimates of exactly zero. More formally, the LASSO limits the sum of absolute partial correlation coefficients which in turn lead to overall shrinkage of estimates and inevitably some become zero. Mathematically, if $\hat{\sigma}$ represents the sample variance–covariance matrix) LASSO aims to estimate the precision matrix by maximizing the penalized likelihood function (with λ_k being the penalty parameter).

$$l(\Theta) = \log \det \Theta - \text{tr}(\hat{\sigma} \Theta) - \lambda_k \sum_{i,j} (|\Theta_{i,j}|)$$

For the purpose of our study, both the significance testing and the graphical LASSO serve as a robustness check for identifying the true network that emerges between Bitcoin owner groups.

3. Data

We consider all data from the Bitcoin blockchain, from 25 February 2012 to 17 July 2017 (1969 days with 1843 observed days), described in detail in Chen et al. (2018). Bitcoin blocks are published approximately every 10 min and contain information about the transaction size, the account ID (anonymous), the participating accounts and the timestamp of the transactions.

The previous information is very useful to understand the time dynamics of volume transactions, but it indicates nothing about the nature of the bitcoin owners who generate the trade. Trying to capture some kind of information on bitcoin traders, we consider the website Blockchain.info provides information about the IP address of the relying party that provides a secure access to the originator of each transaction, and extract from it the approximate geographical provenience of the trader who generates the transaction. To avoid a too large approximation error, we decided to group geographical provenience in a few classes, corresponding to six continental groups: Africa (Af), Asia (As), Europe (Eu), North America (N_A), Oceania (Oc) and South America (S_A). More precisely, the continent of the bitcoin trader is identified from the data in Blockchain.info, comparing its IP address with a dataset of IP address from MaxMind Inc. The approximate location of the transaction origin can be tracked by recording the first node relaying it. We remark that this approach works as long as the running node does not use an anonymizing technology.

We thus have a first grouping of bitcoin owners that roughly correspond to their continent of residence. To further characterize them, for each of the six continental groups we associate to each account IDs according the absolute size of the total transaction amount they generate in the considered time period. We then further group the IDs of each continent according to the deciles of their statistical distribution. The first group, which will be labeled 1 after the continent abbreviation, has the smallest

transactions, corresponding to the 0–10% percentile class, while the tenth group with the largest transactions is labeled 10, corresponding to 90–100% percentile class. The final result is a classification of bitcoin owners in 60 groups: 10 groups per continent.

With this grouping we will investigate our research hypotheses, and search for the bitcoin owners who mostly impact the market. Specifically we will be able to investigate whether large-size Bitcoin owner affect the trade decisions of the others, or whether a specific continent drives the others, in terms of bitcoin trades, or both.

We remark that, although the Bitcoin is the most liquid and largest cryptocurrency, there is sometimes low liquidity in its transactions. Our data show that there are days without a single transaction in Africa, Asia, Oceania and South America, with frequency of low liquidity varying between 1% and 25%. We can overcome the liquidity problem by accumulating the 10 min data to a daily frequency. In any case, this indicates that a further regional grouping, for example by countries, would lead to lack of data for many of them.

For each of our considered groups, our main variable of interest is the volume of transactions, in any given time point. To normalise such data, we consider the logarithm of the transaction volumes. To avoid computational problems, when no transactions in a group arise within a day, we add 1 Satoshi¹ to each transaction. Given the large numbers under consideration, the bias effect of the correction is negligible.

In Figure 1 we illustrate the daily log accumulated transaction sizes over all 10 groups in each continent. The largest transaction sizes appear in Europe and North America, whose dynamic pattern is quite steady. Asia and Oceania are evidently more volatile than Europe and North America, but less volatile than Africa and South America. The descriptive statistics, reported in Table 1, provide further evidence to these findings. Note in particular that Asia, Oceania, Africa and South America have a minimum value of zero, indicating a lack of liquidity in certain time periods.

For deeper insights into the data features of the groups in each continent, the empirical distribution of the log transaction sizes is displayed by means of boxplots in Figure 1. For each continent, the left plot corresponds to the first group, namely the group 1 with the smallest transactions, and the right one to the group 10 with the largest transactions, respectively.

Table 1. Descriptive statistics of the accumulated log transactions of the 6 regions Africa (Af), Asia (As), Europe (Eu), North America (N_A), Oceania (Oc), South America (S_A). Eu and N_A show a related behavior in terms of the descriptive statistics, as so do As and Oc. Also Af and S_A behave related.

	Af	As	Eu	N_A	Oc	S_A
mean	142.25	193.77	232.18	230.45	186.60	155.80
sd	72.84	19.81	11.59	9.18	24.55	62.39
skewness	−1.30	−4.81	−0.86	−1.61	−4.59	−1.91
kurtosis	2.98	44.71	5.27	10.50	34.79	5.12
min	0.00	0.00	162.72	154.25	0.00	0.00
max	222.76	240.14	257.76	254.96	235.36	228.09

From Figure 1, the narrow box width of Europe and North America suggests that these continents are characterised by transaction sizes with low volatility and a few outliers. However for Asia and Oceania the daily transaction sizes are more volatile, and lead to larger center boxes and wider whiskers. South America becomes extreme in the sense of showing even longer whiskers, with transaction sizes varying stronger between groups. Africa follows a very different picture from the other continents: it has the lowest liquidity and a much higher volatility and it shows frequent drops of the transaction volume to 0.

¹ The BTC transactions are reported in Satoshi values, the smallest fraction of a BTC, where 1 BTC = 100,000,000 Satoshi.

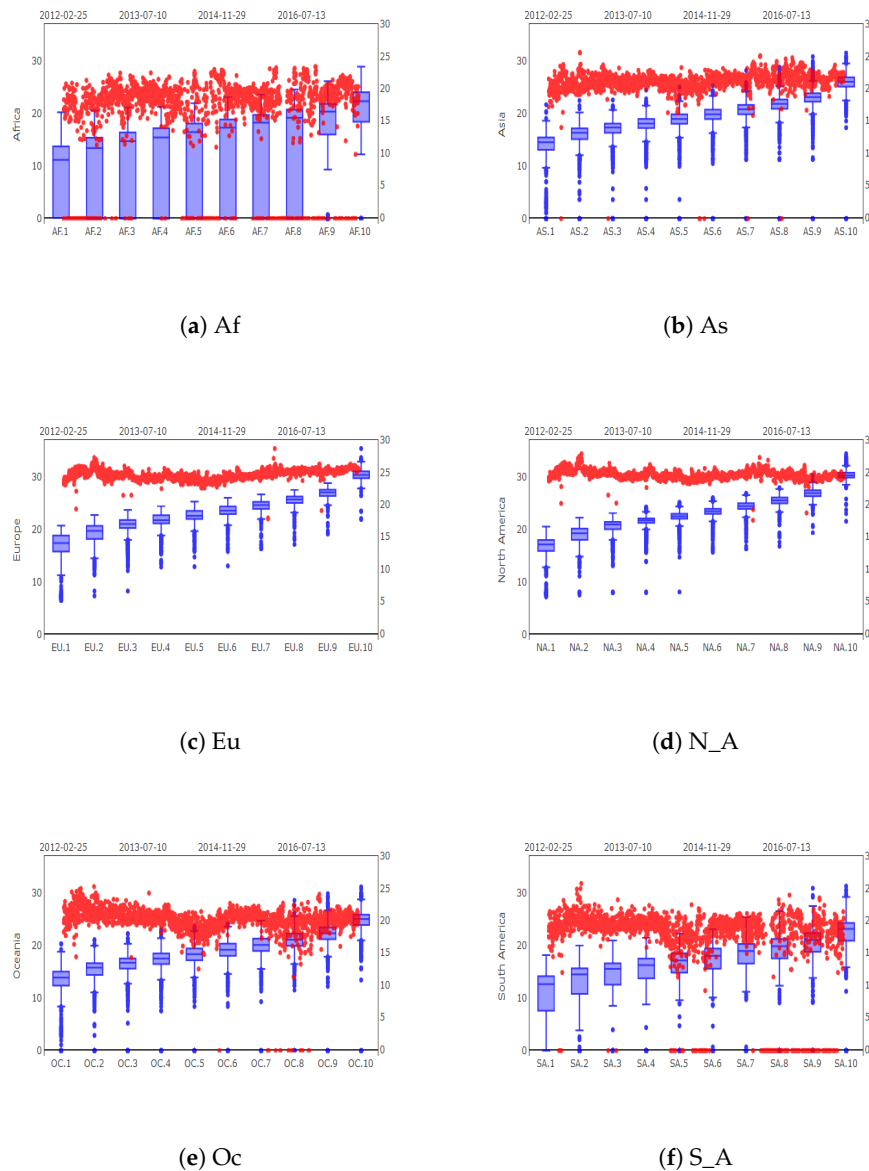


Figure 1. Daily volume transactions (expressed in logarithms) of the 10 groups displayed as boxplots, where the left boxplot represents the first group and the right one the ten groups of the respective continent. The scatter plot displays the accumulated log transaction size of the 10 groups. The time period goes from 25 February 2012 until 17 July 2017 in all 6 continents.

4. Empirical Findings

In this Section we present the results from the application of the proposed model. First we evaluate the model in terms of predictive accuracy, to gauge its validity in the present context; second, we interpret the model results in terms of our research hypotheses, aimed at assessing the dependency patterns among the trading behaviour of different bitcoin traders.

We first consider an unregularised network, whose edges are all present, even when the corresponding partial correlation is very low.

By calculating the partial correlations as specified in (6), we can derive the B_0 matrix and, then, the autoregressive parameters A_1, \dots, A_{p0} . We are thus able to disentangle the time-dependent volume of node i , separately estimating the autoregressive idiosyncratic component and the contemporaneous one, according to Equation (2). Table 2 presents the assessment of the predictive performance of our

model, to understand if the proposed approach is suitable, from a statistical viewpoint. Specifically, we want to investigate whether the inclusion of the contemporaneous component improves predictive accuracy, with respect to a much simpler pure autoregressive model. Table 2 contains the results of the predictive assessment.

Table 2. Comparison between the root mean square errors obtained with our full VAR model and with a model composed by the solely autoregressive component.

Group	RMSE_Full	RMSE_AR	Group	RMSE_Full	RMSE_AR
Africa1	0.1945	0.2052	N_A1	0.2495	0.2500
Africa2	0.1298	0.1315	N_A2	0.4590	0.4613
Africa3	0.1600	0.1584	N_A3	0.5523	0.5596
Africa4	0.1521	0.1538	N_A4	0.3241	0.3631
Africa5	0.1492	0.1460	N_A5	0.8437	0.8530
Africa6	0.1609	0.1538	N_A6	1.2396	1.2653
Africa7	0.1385	0.1419	N_A7	0.9865	0.9951
Africa8	0.1382	0.1371	N_A8	0.8721	0.9041
Africa9	0.1276	0.1250	N_A9	0.6895	0.6962
Africa10	0.0960	0.0979	N_A10	1.2575	1.2698
Asia1	0.2258	0.2286	Oceania1	0.3182	0.3209
Asia2	0.2340	0.2264	Oceania2	0.2447	0.2477
Asia3	0.3148	0.3173	Oceania3	0.3717	0.3655
Asia4	0.3479	0.3432	Oceania4	0.4795	0.4914
Asia5	0.4328	0.4501	Oceania5	0.4909	0.5057
Asia6	0.5425	0.5493	Oceania6	0.5837	0.5782
Asia7	0.6143	0.6064	Oceania7	0.5857	0.5965
Asia8	0.6403	0.6455	Oceania8	0.8265	0.8353
Asia9	0.5294	0.6863	Oceania9	0.3350	0.3255
Asia10	0.5565	0.5623	Oceania10	0.2659	0.2733
Europe1	0.0558	0.0572	S_A1	0.2577	0.2663
Europe2	0.1414	0.1433	S_A2	0.2162	0.2183
Europe3	0.1779	0.1894	S_A3	0.2315	0.2326
Europe4	0.1405	0.1423	S_A4	0.2307	0.2302
Europe5	0.1822	0.1839	S_A5	0.2196	0.2231
Europe6	0.2241	0.2257	S_A6	0.2227	0.2234
Europe7	0.2852	0.2880	S_A7	0.2152	0.2145
Europe8	0.3673	0.3688	S_A8	0.2052	0.2061
Europe9	0.4021	0.4028	S_A9	0.1970	0.1960
Europe10	0.3460	0.3481	S_A10	0.1749	0.1757

From Table 2 note that the proposed model overperforms a pure autoregressive model, as the corresponding root mean squared errors of the one-step ahead predictions are lower in the vast majority of cases. It can be shown that the overall RMSE is equal to about 0.37 for the proposed model, against 0.42 for the autoregressive one, further confirming its superiority.

We now move towards the interpretation of the results that can be drawn from our model and, specifically, from the partial correlations (Equation (6)). In Figure 2, each node represents one of the 60 groups of traders and each present edge indicate that two traders are dependent on each other, in terms of their transactions (conditionally on all the others). Differently, when an edge is missing, the corresponding traders behave independently of each other (conditionally on all the others). Each edge is associated with a weight, which corresponds to a partial correlation coefficient. The size of each edge in Figure 2 is proportional to such weight. On the other hand, the coloring of an edge between two nodes indicates the sign of the partial correlation coefficient: green highlights a positive partial correlation and red a negative partial correlation.

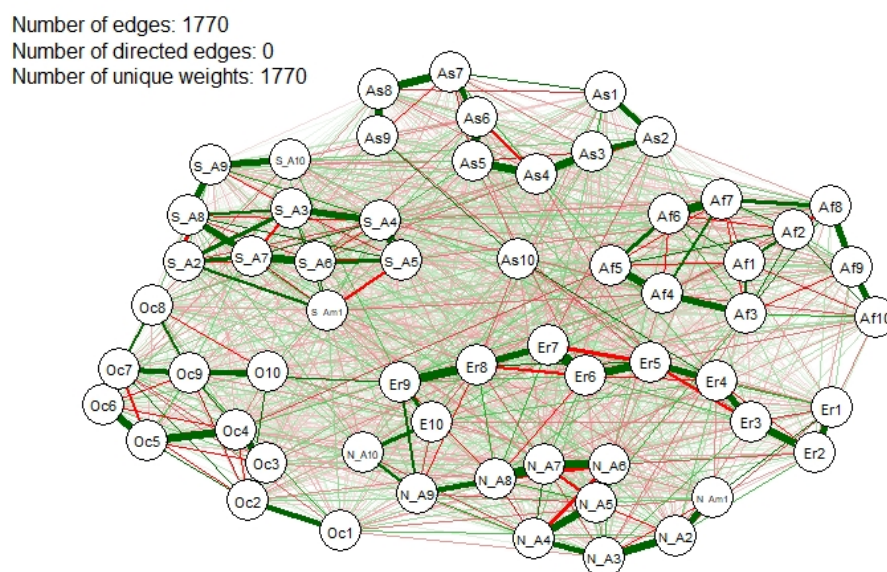


Figure 2. Unregularized Partial Correlation Network.

What we can observe from the network that emerges from Figure 2 is that there exist many interconnections between Bitcoin groups of users. Precisely, the summary statistics provided in the upper left corner of Figure 2 indicates that the network contains a total of 1770 non-zero links between groups. Although the graph is difficult to interpret, some clusters can be identified. We can see about five clusters which in most part correspond to the continents, with the exception of Europe and North America which are placed in the same cluster, suggesting that there exist strong dependence between the traders of the two continents. This is something that we expected to see due to the economic and political similarities among the two regions, as well as on their news sharing.

Note also that the groups representing the larger traders in Europe and North America - N_A10, N_A9, Eu10, Eu9 - show stronger positive connections than other groups. This may be explained by the fact that these groups have a comparable size of transactions, which come from a similar set of information, which induce them to behave similarly. If we match this result with that in Figure 1, which indicates the relatively larger volumes of transactions coming from these groups, we obtain a clear indication that these are the groups which can mostly impact the market. Note also that there exists a strong positive link between Oc10 and Eu9, and not between Oc9 and Eu9. This is consistent with our previous finding: the transaction volumes of Oc10 are more comparable in their size to Eu9, rather than to Eu10 (see Figure 1) and, therefore, they act similarly.

As mentioned previously, in unregularized correlation networks some edges may present but may not be statistically significant. In the graphical representation, such situations will be visualized as very weak connections in the network. To prevent this and to correctly identify the significant associations between Bitcoin groups, a crucial step is to impose restrictions that will limit (or eliminate) the occurrence of spurious edges. One way to achieve this is by testing the statistical significance of partial correlations.

Figure 3 presents the same network containing only links that are found statistically significant at both 5% and 1% level of significance.

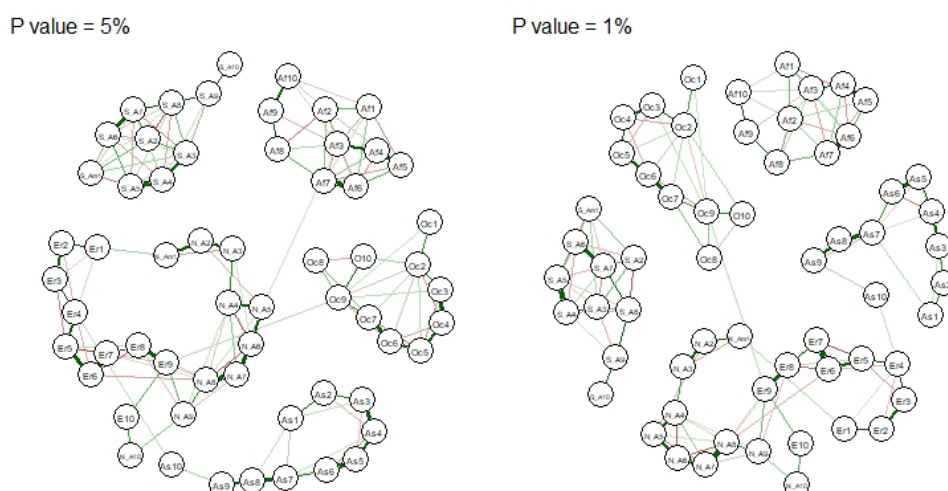


Figure 3. Regularized partial correlation networks (without edges that are not significant).

Figure 3 shows that the structure of the network does not change significantly if we impose different levels of significance. What we observe from the graphs is that the majority of links that were present in the unregularized network have disappeared, reducing the total number of links from 1770 to 146 and 137, respectively. Interesting, even though a significant portion of the links were removed, the clustering of nodes remains the same as in Figure 2. Specifically, we see the formation of clusters equivalent to the continents and we also see significant interconnection between traders in Europe and North America. Furthermore, we also see a statistically significant positive correlation Oceania's top group and Europe's and between Asia's top group and Europe's.

To further confirm our findings, we perform a further robustness check through the application of the graphical LASSO. As discussed previously, LASSO is a very popular method for eliminating spurious links. Figures 4 and 5 represent the networks that emerge by the applying graphical LASSO with different smoothness parameters λ . We remark that, unlike the classical LASSO, in the graphical approach the choice of λ cannot be done based on cross-validation as it represents a completely unsupervised process. As we are mainly interested in assessing the robustness of the results, we consider four alternative values for λ , and see whether what found in Figure 3 changes.

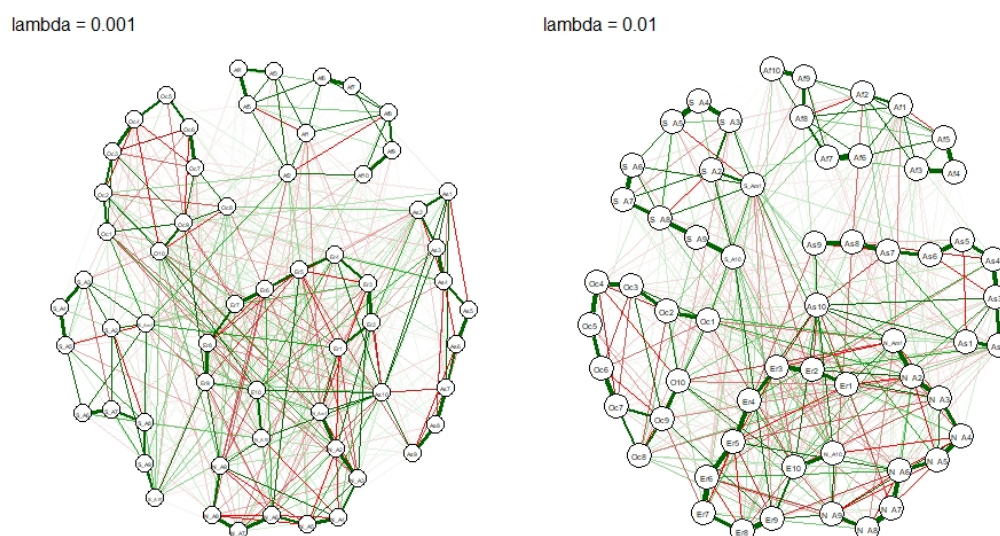


Figure 4. GLASSO partial correlation networks [varying lambda], 1/2.

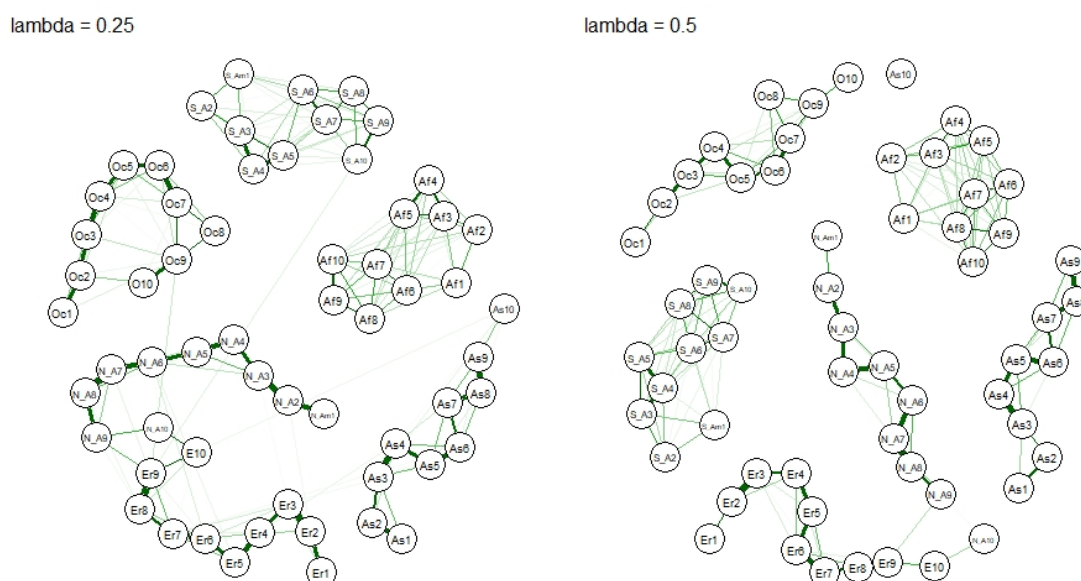


Figure 5. GLASSO partial correlation networks [varying lambda], 2/2.

From Figures 4 and 5, the changing λ does change the structure of the network, but the underlying clusters remain the same, thus confirming the close interconnection between Europe and North America, as well as those between top traders in Oceania and Europe.

A closer inspection of Figure 4, reveals frequent linkages between European and North American nodes, which is in line with the previous observations. Positive linkages appear more often inside each continent, compared to negative ones. On the other hand negative and positive edges appear frequently between two continents (see Table 3). The largest two groups in both continents share strong links with each other, confirming that they probably share a common information set. Interestingly the largest trader group from Asia, AS10, has multiple positive edges to several groups in Europe and North America. Considering that most bitcoin mining farms are based in Asia, and especially in China, it follows that a large amount of capital is acquired and, therefore, traded, from Asia with the rest of the world. Last, note that the largest volume trading groups from Oceania and South America also share links with each other and with the larger Western-World groups. This observation leads to the conclusion that the large traders around the world are somewhat connected, possibly communicating with each other. On the other hand smaller groups, which have less information, shows less connections around the world.

Table 3. Count of links between and within North America and Europe.

	Lambda 0.001		Lambda 0.01	
	Positive	Negative	Positive	Negative
Within Europe	17	14	17	13
Within North America	21	13	19	13
Between Europe and North America	48	53	45	48

Figure 5 shows what happens when we increase the penalty level to $\lambda = 0.25$. Most edges vanish, but the previously found connections persists. Still the largest trader groups from Europe and North America remain connected, while the edges from Oc9, S_A10 and As10 persist to stay connected with them. The connection goes via the largest groups in Europe, namely Eu9 and Eu10. Other persisting edges exist between the smaller groups from Asia and Europe, yet with small magnitude. Within the continents many edges are not affected by the penalty, hence emphasize the importance of the regional connectedness. Finally, when increasing the penalty parameter to $\lambda = 0.5$, most cross-continent

edges are ruled out, except for the ones between the largest groups in Europe and North America. The remaining edges only appear within the continents.

To further establish the robustness of the results to the varying value of λ , Table 4 compares some centrality values, averaged over the whole network, under the four considered values of λ .

Table 4. Average centralities across different lambda parameters.

	$\lambda = 0.001$	$\lambda = 0.01$	$\lambda = 0.25$	$\lambda = 0.5$
Average degree	1.189206937	1.157479	0.855028	0.663931
Average betweenness	270.5666667	288.5667	269	39.3
Average closeness	0.000448235	0.000428	0	0

From Table 4 note that, consistently with our previous findings, by increasing the parameter λ the average centrality decreases, according to degree, betweenness and closeness. Regardless of this, our main conclusions remain stable.

To summarise, our empirical findings give an answer to our research proposition: which are the group of traders that mostly affect the bitcoin markets? These groups were found among the top two classes of traders in North America and Europe, strongly and positively connected to each other. These traders are linked to the others, affecting their behaviours. In particular, they are especially linked with the top traders from Oceania and South America. In addition, top traders from Asia, and especially larger ones, are highly linked to the others, likely as a result of their mining activity.

5. Conclusions

In the paper, we proposed a model that explains the dynamics of Bitcoin trading volumes, based on a correlation network VAR process that models the interconnections between different groups of traders.

Our main methodological contribution consists of the introduction of partial correlations and correlation networks into VAR models. This allows describing the correlation patterns between trading volumes and to disentangle the autoregressive component of volumes from its contemporaneous part. The introduction of VAR correlation networks also allows building a volume predictive model that leverages the information contained in the correlation patterns.

Our main financial findings show that trading volumes are highly correlated within geographical regions. Groups of traders with high transaction volumes over all continents covary in the network model, leading to the conclusion this groups share a mutual information set. The results are robust over various penalized network models. This result may have different economical explanations, such as a common behaviour, a common time-zone, similar institutional and legal contexts.

Our results also contribute to the identification of group of bitcoin traders that are the most likely influencers of the market. These are found to high volume traders, especially from North America, Europe, and Asia. These results are in line with the expectation that trading follows the news sharing patterns and the major Bitcoin mining localization patterns.

The proposed model can be very useful for policy makers and regulators. It can be used to predict “regular” trading volumes and, therefore, identify anomalies. Our empirical findings show that the proposed model is able to predict trading volumes with an error that is lower than that of a pure autoregressive model.

Our result suggests that policy makers and regulators, interested in preserving the integrity of bitcoin markets, should also pay particular attention to the transactions coming from large volume traders, and especially of those from America, Europe and Asia, which have the potential to disrupt the market.

The main weakness of this work is related to the available sample. It refers to a specific cryptoasset, the bitcoin; it relates to a specific period of time and is taken directly from blockchain transactions, rather than from market exchanges. These limitations derive from the proprietary nature of the data

that was made available to us. However, we believe that our model is rather general, and can be easily extended on a different database. This in particular to deal with transactions that take place on crypto exchanges, more frequent than those taking place on the blockchain, considered here. Further work may concern acquiring data on the electronic identity of the traders, to investigate the reason of “regional” behaviours, as also discussed in [Tasca et al. \(2018\)](#) and [Foley et al. \(2019\)](#).

From a methodological viewpoint, it may be worth considering extending correlation network models to become time dependent, although this requires acquiring data with a higher frequency. In addition, it may be worth considering an extension of the model that accounts for exogenous factors, such as regulatory interventions, transaction fees, sentiment and media coverage. This may require an event-based analysis, aimed at understanding not only trading patterns, but also what may originate them. To achieve this task our work could be extended with Bayesian network models, following [Giudici et al. \(2014\)](#), [Giudici and Bilotta \(2014\)](#) and [Cerchiello and Giudici \(2016\)](#).

Author Contributions: All four authors have contributed to the paper and, in particular, to its conceptualization, methodology, software, data curation, validation, writing, review and editing. The paper work has been coordinated by the corresponding author. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no specific external funding.

Acknowledgments: We acknowledge useful comments and suggestions from the participants at the workshops where the paper was presented. We also acknowledge very useful comments and suggestions from the four referees that have commented the paper very thoroughly. The comments have helped us to substantially revise the paper. This research has received funding from the European Union’s Horizon 2020 research and innovation program “FIN-TECH: A Financial supervision and Technology compliance training programme” under the grant agreement No 825215 (Topic: ICT-35-2018, Type of action: CSA). We also gratefully acknowledge the financial support of Singapore Ministry of Education Academic Research Fund Tier 1 at National University of Singapore.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Acharya, Viral, Robert Engle, and Matthew Richardson. 2012. Capital shortfall: A new approach to ranking and regulating systemic risks. *American Economic Review: Papers and Proceedings* 102: 59–64. [\[CrossRef\]](#)
- Acharya, Viral, Lasse Pedersen, Thomas Philippon, and Matthew Richardson. 2016. Measuring systemic risk. *Review of Financial Studies* 30: 2–47. [\[CrossRef\]](#)
- Adrian, Tobias, and Markus Brunnermeier. 2015. Covar. *American Economic Review: Papers and Proceedings* 106: 1705–41. [\[CrossRef\]](#)
- Ahelegbey, Daniel, Monica Billio, and Roberto Casarin. 2016. Bayesian graphical models for structural vector autoregressive processes. *Journal of Applied Econometrics* 31: 357–86. [\[CrossRef\]](#)
- Battiston, Srefano, Domenico Delli Gatti, Mauro Gallegati, Bruce Greenwald, and Joseph Stiglitz. 2012. Liasons dangereuses: Increasing connectivity risk sharing, and systemic risk. *Journal of Economic Dynamics and Control* 36: 1121–41. [\[CrossRef\]](#)
- Bianchi, Daniele. 2019. Cryptocurrencies as an asset class? an empirical assessment. *Journal of Alternative Investments*. [\[CrossRef\]](#)
- Bianchi, Daniele, and Alexander Dickerson. 2019. Trading volumes in cryptocurrency markets *WBS Finance Group Research Paper*. [\[CrossRef\]](#)
- Billio, Monica, Mila Getmansky, Andrew Lo, and Liorana Pelizzon. 2012. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics* 104: 535–59. [\[CrossRef\]](#)
- Bohte, Rick, and Luca Rossini. 2019. Comparing the forecasting of cryptocurrencies by bayesian time-varying volatility models. *Journal of Risk and Financial Management* 12: 150. [\[CrossRef\]](#)
- Bouoiyour, Jamal, Refk Selmi, Aviral Tiwari, and Olayeni Olalolu. 2016. What drives Bitcoin price? *Economics Bulletin* 36: 843–50.
- Bouri, Elie, Chi Keung Lau, Brian Lucey, and David Roubaud. 2019. Trading volume and the predictability of return and volatility in the cryptocurrency market. *Finance Research Letters* 29: 340–46. [\[CrossRef\]](#)
- Catania, Leopoldo, Stefano Grassi, and Francesco Ravazzolo. 2019. Forecasting cryptocurrencies under model and parameter instability. *International Journal of Forecasting* 35: 485–501. [\[CrossRef\]](#)

- Cerchiello, Paola, and Paolo Giudici. 2016. Big data analysis for financial risk management. *Journal of Big Data* 3: 1–18. [CrossRef]
- Chen, Ying, Simon Trimborn, and Jiejie Zhang. 2018. Discover Regional and Size Effects in Global Bitcoin Blockchain Via Sparse-Group Network Autoregressive Modeling. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3245031 (accessed on 1 October 2019). [CrossRef]
- Diebold, Francis, and Kamil Yilmaz. 2014. On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics* 182: 119–34. [CrossRef]
- Elendner, Hermann, Simon Trimborn, Bobby Ong, and Teik Ming Lee. 2017. The Cross-Section of Crypto-Currencies as Financial Assets: Investing in crypto-currencies beyond bitcoin. In *Handbook of Blockchain, Digital Finance and Inclusion: Cryptocurrency, FinTech, InsurTech, and Regulation* 1st ed. Edited by D. Lee Kuo Chuen and R. Deng. Amsterdam: Elsevier, vol. 1, pp. 145–73.
- Foley, Sean, Jonathan Karlsen, and Talis Putnins. 2019. Sex, Drugs and Bitcoin. how much illegal activity is financed through cryptocurrencies *Review of Financial Studies* 32: 1798–853.
- Giudici, Paolo, and Iman Abu-Hashish. 2019. What determines bitcoin exchange prices? a network var approach. *Finance Research Letters* 28: 309–18. [CrossRef]
- Giudici, Paolo, and Annalisa Bilotta. 2004. Modelling operational losses: A Bayesian approach *Quality and Reliability Engineering* 20: 407–17. [CrossRef]
- Giudici, Paolo, Maura Mezzetti, and Pietro Muliere. 2003. Mixtures of products of Dirichlet process for variable selection in survival analysis. *Journal of Statistical Planning and Inference* 111: 101–15. [CrossRef]
- Giudici, Paolo, and Paolo Pagnottoni. 2019a. High frequency price change spillovers in bitcoin exchange markets. *Risks* 7: 111. [CrossRef]
- Giudici, Paolo, and Paolo Pagnottoni. 2019b. Vector error correction models to measure connectedness of bitcoin exchange markets. *Applied Stochastic Models in Business and Industry*, in press. [CrossRef]
- Giudici, Paolo, and Gloria Polinesi. 2019. Crypto price discovery through correlation networks. *Annals of Operations Research*. [CrossRef]
- Giudici, Paolo, and Alessandro Spelta. 2016. Graphical network models for international financial flows. *Journal of Business and Economic Statistics* 34: 128–38. [CrossRef]
- Ji, Qiang, Elie Bouri, Chi Keung Lau, and David Roubaud. 2019. Dynamic connectedness and integration in cryptocurrency markets. *International Review of Financial Analysis* 63: 257–72. [CrossRef]
- Lauritzen, Steffen. 1996. *Graphical Models*. Oxford: Oxford University Press.
- Lorenz, Jan, Stefano Battiston, and Frank Schweitzer. 2009. Systemic risk in a unifying framework for cascading processes on networks. *The European Physical Journal B—Condensed Matter and Complex Systems* 71: 441–60. [CrossRef]
- Makarov, Igor, and Antoninette Schoar. 2019. Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics*. [CrossRef]
- Tasca, Paolo, Shaowen Liu, and Adam Hayes. 2018. The evolution of bitcoin economy: Extracting and analyzing the network of payment relationship. *The Journal of Risk Finance* 19: 94–126. [CrossRef]
- Whittaker, Joe. 1990. *Graphical Models in Applied Multivariate Statistics*. Chichester: John Wiley and Sons.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).